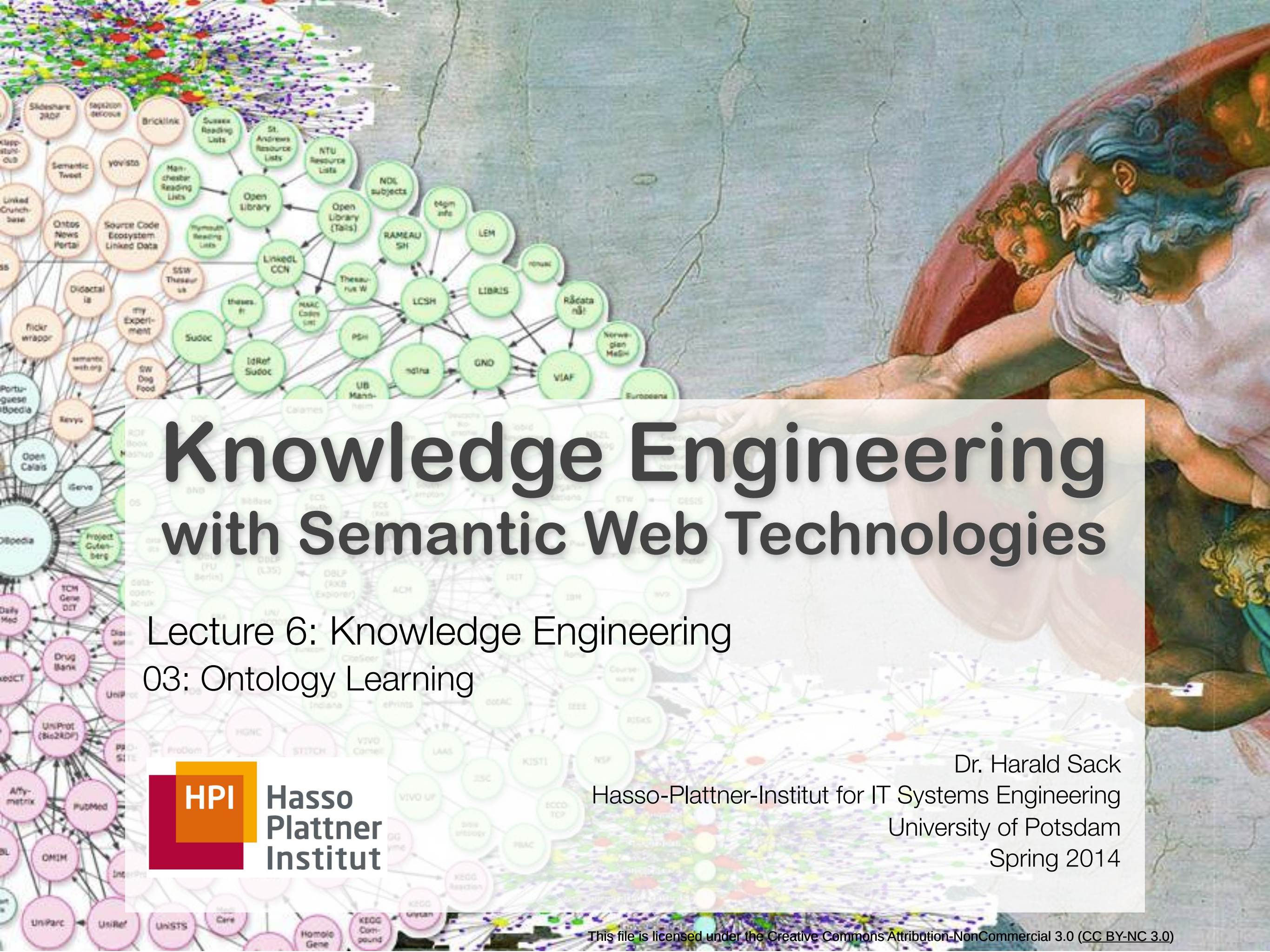




Knowledge Engineering with Semantic Web Technologies

Lecture 6: Knowledge Engineering
03: Ontology Learning

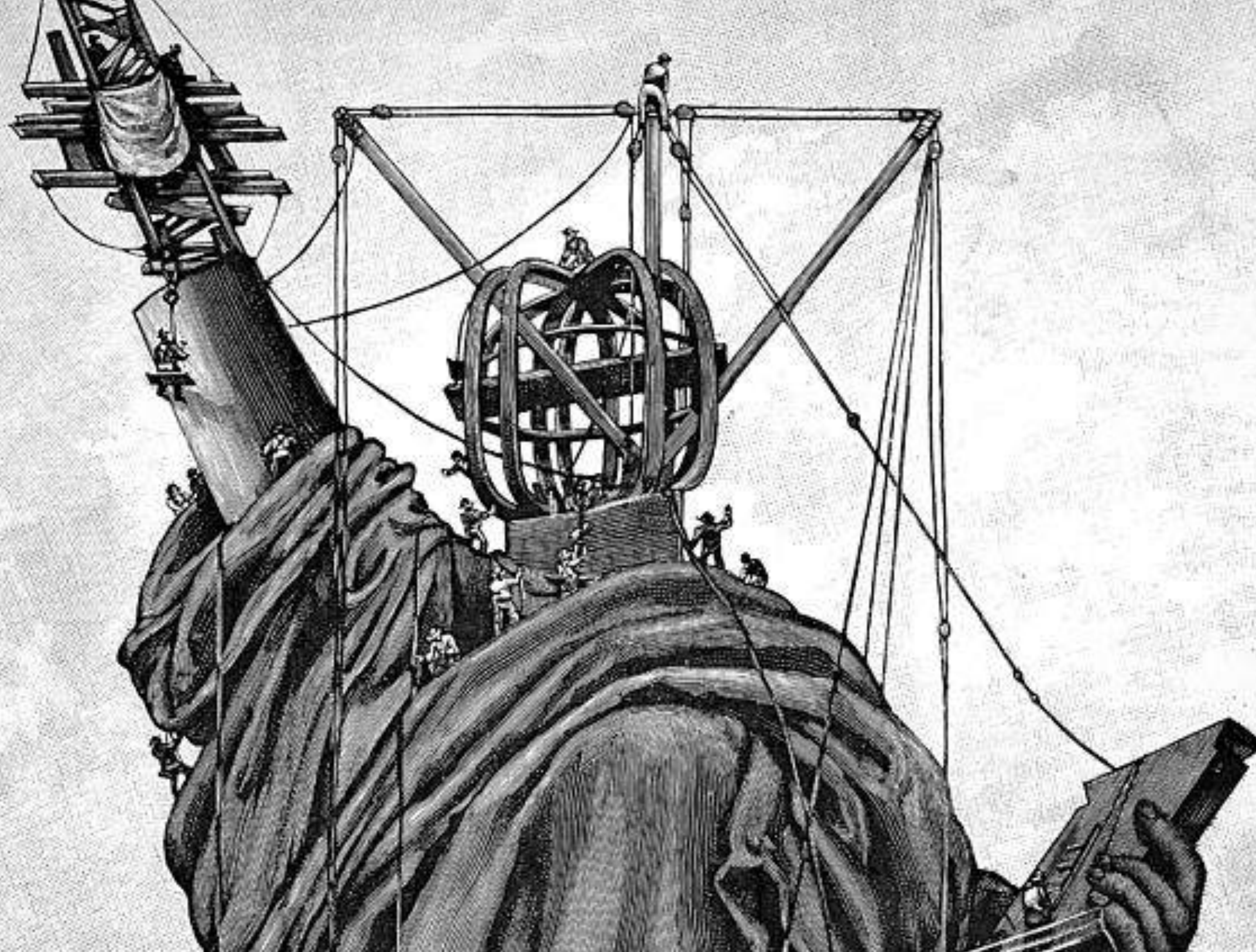


Dr. Harald Sack
Hasso-Plattner-Institut for IT Systems Engineering
University of Potsdam
Spring 2014



Lecture 6: Knowledge Engineering

Open HPI Course: Knowledge Engineering with Semantic Web Technologies



03 - Ontology Learning

Open HPI - Course: Knowledge Engineering with Semantic Web Technologies
Lecture 6: Knowledge Engineering

Ontology Learning

- Ontology Design is very expensive wrt. time and resources
 - *can we automate the process or at least some parts?*
- Ontologies can be „learned“ automatically
- **Ontology Learning** defines a set of methods and techniques
 - for **fundamental development** of new ontologies
 - for **extension or adaption** of already existing ontologies
 - in a (partly) automated way from various resources.
- also referred to as *Ontology Generation, Ontology Mining, or Ontology Extraction*
- Automatization requires help from
 - Natural Language Processing (NLP)
 - Data Mining
 - Machine Learning techniques (ML)

Ontology Learning

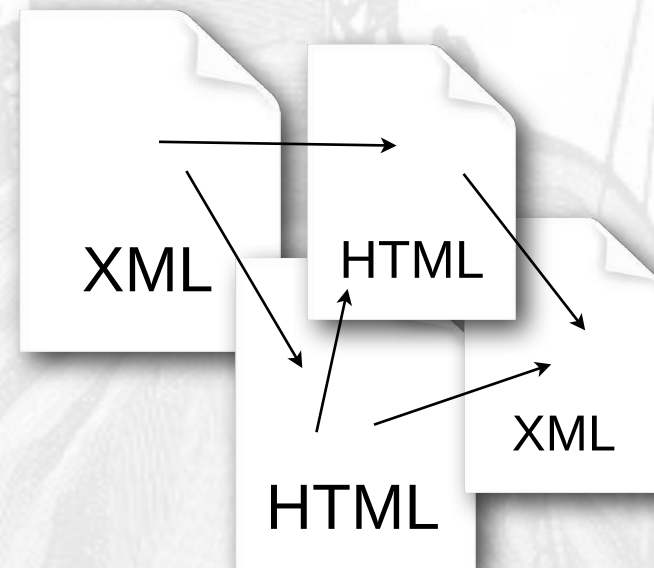
- Fundamental **(input) data** for ontology learning:

- **Structured Data**

A small image of a historical data table with multiple columns and rows, representing structured data.

Machine Learning

- **Semi-structured Data**



**Natural Language Processing
+**

- **Unstructured Data**

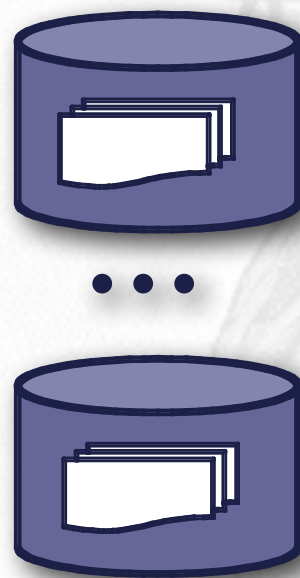


Machine Learning

Ontology Learning - Basic Approach (I)

6

document corpus



(1) Term
extraction



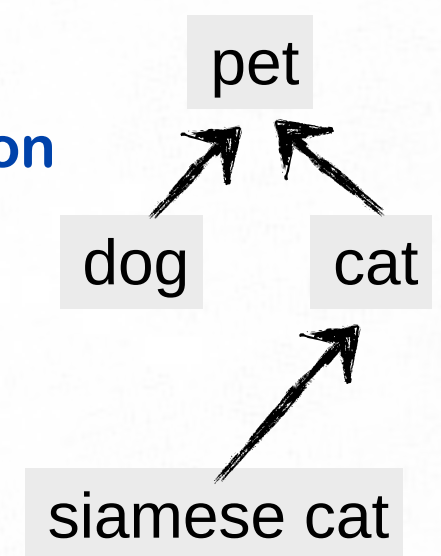
terms

<dog> <dogs>
<cat>
<siamese cat>

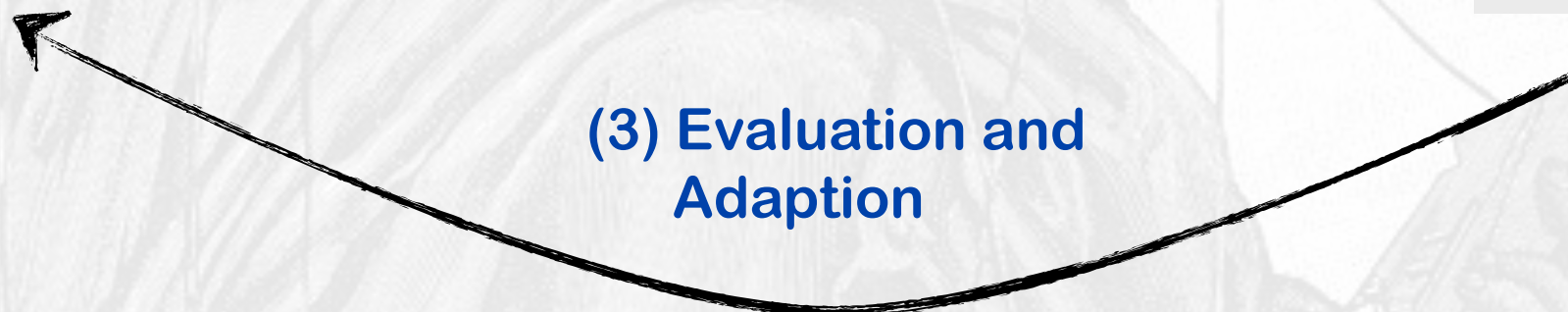
(2) Conceptualization



ontology



(3) Evaluation and
Adaption



Term extractions: requires linguistic processing (NLP) to identify important noun phrases and their internal semantic structure

Terms: linguistic realizations of domain specific concepts

Concepts: clusters of semantically related terms

• Natural Language Processing:

(1) Tokenizer / Sentence Splitter

(2) Morphological Analysis

- Stemming
- Lemmatizer

(3) POS-Tagger

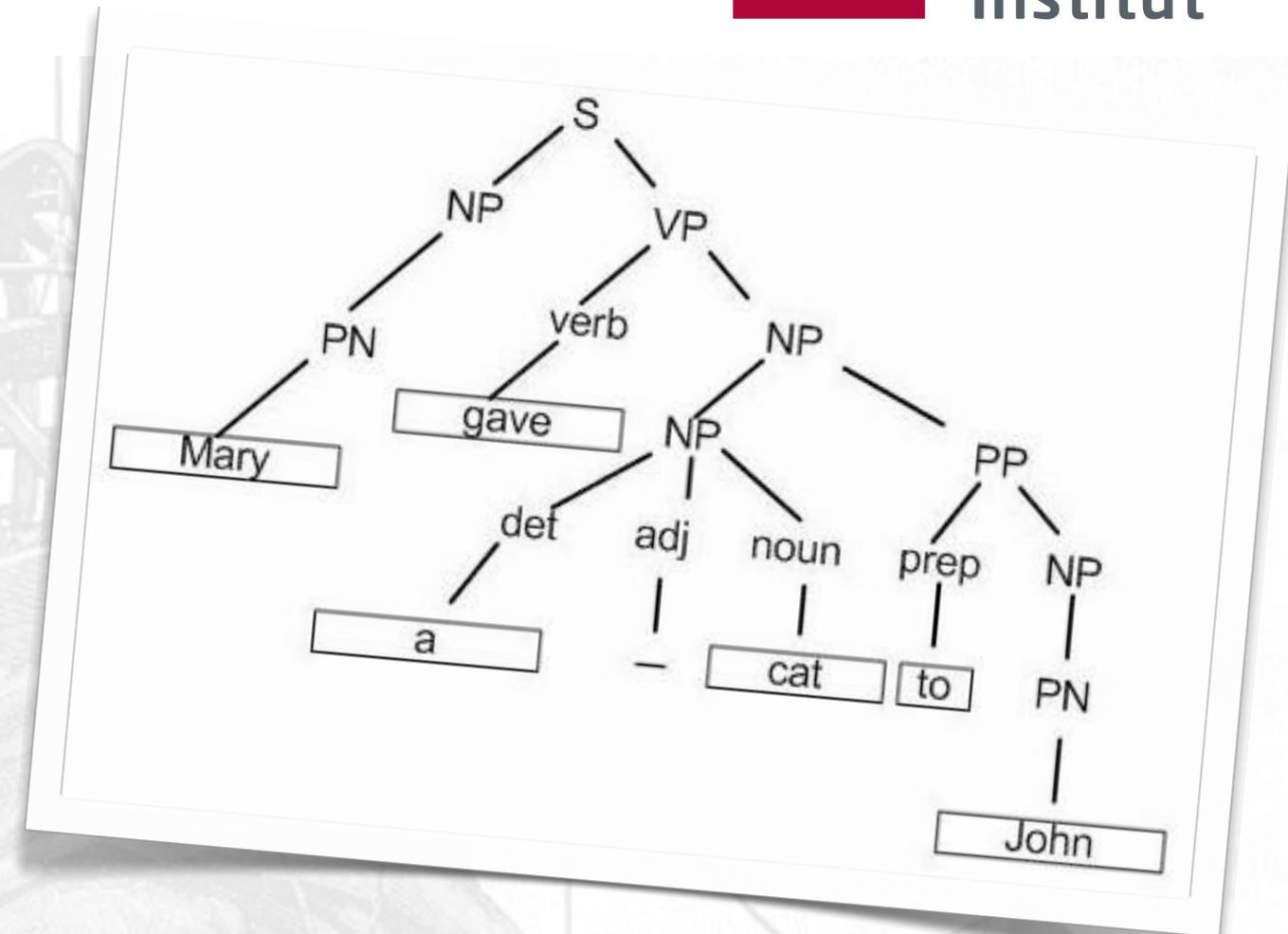
- Syntactic categories
(verb, noun, preposition, etc...)

(4) Regular Expression Matching

(5) Chunks

- Detection of larger coherent structures in sentences

(6) Syntactic Parser



Ontology Learning - Basic Approach (III)

8 • Machine Learning:

(1) Association Rule Discovery

- Discovery of interesting associations between terms

(2) (Hierarchical) Clustering

- unsupervised learning, in particular clustering of terms

(3) Classification

- Classification of new concepts into an existing hierarchy
- e.g. with Support Vector Machines (SVM), Naive Bayes, knn

(4) Inductive Logic Programming

- Induction of rules from data, i.e. discovery of new concepts from extensional data

(5) Conceptual Clustering

- Formal Concept Analysis, learning concepts and concept hierarchies

Ontology Learning - Layer Cake

$\forall x((\text{country}(x) \rightarrow \exists y \text{ capitalOf}(y,x)) \wedge (\forall z (\text{capitalOf}(z,x) \rightarrow y=z)))$

General Axioms

$\text{river} \sqcap \text{mountain} = \emptyset$

Axiomatic Schemata

$\text{capitalOf} \sqsubseteq_R \text{locatedIn}$

Hierarchy of Relations

$\text{flowThrough}(\text{dom:river}, \text{range:GeoEntity})$

Relations

$\text{capital} \sqsubseteq_c \text{city}, \text{city} \sqsubseteq_c \text{InhabitedGeoEntity}$

Hierarchy of Concepts

$c := \text{country} := \langle \text{description}(c), \text{uri}(c) \rangle$

Concept Description

$\{\text{country}, \text{nation}, \text{Land}\}$

Multilingual Synonyms

$\text{river}, \text{country}, \text{nation}, \text{city}, \text{capital}, \dots$

Terms

Ontology Learning Tasks

- which tasks from ontology development can be automated?

Ontology Learning Tasks

- Ontology creation
- Ontology schema extraction
- Extraction of ontology instances
- Ontology integration and navigation
- Ontology update
- Ontology enrichment

Ontology Learning Tasks

11

- **Ontology Creation**

- Design of an ontology from the scratch by a team of experts
- *Maschine Learning* (ML) supports the experts during the design phase by
 - Suggestions of well suited relations among concepts
 - Integrity / consistency checking of the designed ontology
 - ...

Ontology Learning Tasks

12

- **Ontology Schema Extraction**
 - Extraction of ontology schemata from heterogeneous documents, such as e.g. web documents or text documents
 - *Machine Learning* uses **input data** and **meta ontology** to create fully-fledged domain ontologies (with the help of human experts)

Ontology Learning Tasks

- **Extraction of Ontology Instances (Populate Ontologies)**
 - Extraction of ontology instances from semi-structured data or unstructured data to populate already existing ontology schemata with individuals
 - applies technologies from *Information Retrieval* and *Data Mining*

Ontology Learning Tasks

14

- **Ontology Integration and Navigation**
 - Reconstruction of existing knowledge bases and navigation in existing knowledge bases,
 - as e.g. translation of an existing knowledge base from FOL to OWL 2 DL

Ontology Learning Tasks

15

- **Ontology Update**

- Extension, reconstruction and adaption of already existing ontologies, as e.g. adation of an ontology to a changed domain
- relates to parts of ontologies that have been created in the way that they can be changed

Ontology Learning Tasks

16

- **Ontology Enrichment**

- (also *Ontology tuning*) relates to automated update of smaller parts of existing ontologies
- doesn't changes concepts and relations, but refines them (more precise)
- in difference to ontology update only parts of the ontology are considered that usually shouldn't be changed



04 - Ontology Alignment

Open HPI - Course: Knowledge Engineering with Semantic Web Technologies
Lecture 6: Knowledge Engineering